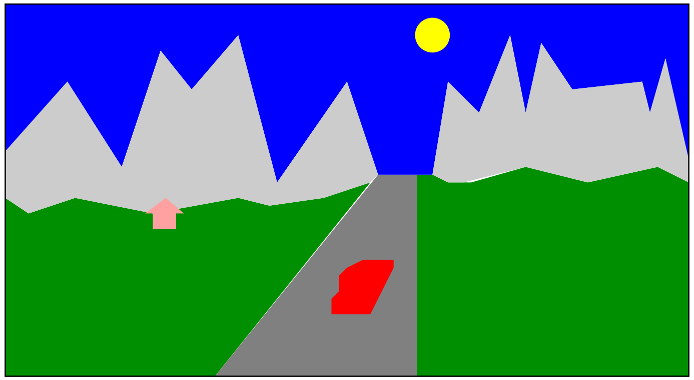
12 Markov Random Fields

- können als Erweiterungen eindimensionaler Markov-Ketten aufgefaßt werden
- ursprüngliche Motivation zur Betrachtung solcher Prozesse stammt aus der statistischen Physik:

  Ising-Modellen zur wahrscheinlichkeitstheoretischen Modellierung von ferro-magnetischem Material

- Einsatz zur Klassifikation für die Berücksichtigung von Kontextabhängigkeiten:

  mehrere benachbarte Entitäten werden gemeinsam, kontextabhängig betrachtet



- klassifiziere die Regionen im Bild
- als: Himmel, Sonne, Auto, Wiese, Meer, . . .
- berücksichtige Kontext, z.B. Sonne am Himmel, nicht im Meer

12.1 Definitionen und Eigenschaften

- Repräsentation eines gegebenen Problems mit Menge von Entitäten oder **Stellen** (*sites*)
$$\mathcal{S} = \{s_1, \ldots, s_N\}$$
z.B. ein Dipol des Ferromagneten oder eine zu klassifizierende Region
- jeder Stelle soll eine **Marke** (*label*) zugewiesen werden:
$$\mathcal{L} = \{l_1, \ldots, l_Q\}$$
diese beschreibt den Zustand (up, down) oder die Klasse (Himmel, Sonne, . . .)
- jeder Stelle s_i wird eine Zufallsvaribale X_i zugeordnet und kann Werte aus $\mathcal{L}$ annehmen
- die Menge $\mathcal{X} = \{X_1, \ldots, X_N\}$ dieser Zufallsvariablen wird ein **Random Field** genannt
- Menge aller möglichen Konfigurationen (X_i nimmt Wert ω_i an):
$$\Omega = \{\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) | \omega_i \in \mathcal{L}\}$$

- Kontextbezüge durch Nachbarschaftssystem modelliert:
  - betrachte die Stellen als Knoten eines ungerichteten Graphen
  - Kantenmenge: $E \subset \{\{s_k, s_l\} | s_k, s_l \in \mathcal{S}\}$
    z.B. angrenzende Regionen, 4er-Nachbarschaft von Bildpunkten
  - die Kanten definieren ein Nachbarschaftssystem $\mathcal{N} = \{N_i | i = 1, \ldots, N\}$

$$\begin{aligned} s_i &\notin N_i \\ s_j &\in N_i, \text{ genau dann wenn } \{s_i, s_j\} \in E \end{aligned}$$

- (für später) betrachte Cliquen: Teilmengen der Knotenmenge, deren Elemente paarweise im Graphen benachbart sind

  alle Cliquen $\mathcal{C}$ werden entsprechend ihrer Mächtigkeiten zusammengefaßt:

$$\begin{aligned} \mathcal{C}_1 &:= \{\{s_i\} \mid s_i \in \mathcal{S}\} \\ \mathcal{C}_2 &:= \{\{s_i, s_j\} \mid s_i, s_j \in \mathcal{S}, \{s_i, s_j\} \in E\} \\ \mathcal{C}_3 &:= \{\{s_i, s_j, s_k\} \mid s_i, s_j, s_k \in \mathcal{S}, \{s_i, s_j\}, \{s_i, s_k\}, \{s_j, s_k\}, \in E\} \end{aligned}$$

- ein Random Field heißt **Markov Random Field** gdw.

$$P(\mathcal{X} = \boldsymbol{\omega}) \; > \; 0 \;\; , \;\; \forall \omega \in \Omega \qquad (12.1)$$

$$P(X_i = \omega_i | X_j = \omega_j, \forall s_j \in \mathcal{S} \wedge i \neq j) \; = \; P(X_i = \omega_i | X_j = \omega_j, s_j \in N_i) \qquad (12.2)$$

  – erste Bedingung ist technischer Natur
  – zweite Bedingung beschreibt die Markov-Eigenschaft mit der lokalen Charakteristik der Kontextabhängigkeit

- die bedingten Wahrscheinlichkeiten bestimmen eindeutig die Verbundwahrscheinlichkeit $P(\mathcal{X} = \boldsymbol{\omega})$, aber i.A. schwierig
  – Funktionen zur lokalen Charakterisierung zu finden, die auch tatsächlich bedingte Wahrscheinlichkeiten darstellen
  – konkrete Berechung

- $\Rightarrow$ alternative Beschreibung über Gibbs-Verteilung
  ein Random Field $\mathcal{X}$ mit dem Nachbarschaftssystem $\mathcal{N}$ wird ein **Gibbs Random Field** genannt, wenn seine Konfigurationen einer **Gibbs-Verteilung** genügen:

$$P(\mathcal{X} = \boldsymbol{\omega}) = \frac{e^{-\frac{1}{T}U(\boldsymbol{\omega})}}{Z} \tag{12.3}$$

  - $U(\boldsymbol{\omega})$ ist die Energie der Konfiguration $\boldsymbol{\omega}$: $U(\boldsymbol{\omega}) = \sum_{c \in \mathcal{C}} V_c(\boldsymbol{\omega})$

    (Summe der **Clique-Potentiale** $V_c(\boldsymbol{\omega})$)
  - Z ist die **partition function**, dient der Normierung $Z := \sum_{\boldsymbol{\omega} \in \Omega} e^{-\frac{1}{T}U(\boldsymbol{\omega})}$
  - die Konstante T heißt in Anlehnung an physikalische Vorgänge Temperatur, kontrolliert Glattheit der Energiefunktion

- **Hammersley-Clifford** Theorem
  beweist Äquivalenz von MRFs und Gibbs-Verteilungen:

  ein Random Field $\mathcal{X}$ mit einem Nachbarschaftssystem $\mathcal{N}$ erfüllt für jede Defintion der Clique-Potentiale die Markov-Eigenschaft

- zur Klassifikation interessiert nicht die a priori Wahrscheinlichkeit einer Konfiguration, sondern

  die a posterirori Wahrscheinlichkeit bei vorliegenden Daten $\mathbf{d}$ (Beobachtungen):

$$\begin{aligned} P(\boldsymbol{\omega}|\mathbf{d}) &= \frac{P(\mathbf{d}|\boldsymbol{\omega})\ P(\boldsymbol{\omega})}{P(\mathbf{d})} \\ &= \frac{1}{P(\mathbf{d})}\ \frac{e^{-\frac{1}{T}U(\boldsymbol{\omega})}}{Z}\ P(\mathbf{d}|\boldsymbol{\omega}) \\ &= \frac{1}{P(\mathbf{d})}\ \frac{e^{-\frac{1}{T}U(\boldsymbol{\omega})}\ e^{\ln P(\mathbf{d}|\boldsymbol{\omega})}}{Z} \\ &= \frac{1}{P(\mathbf{d})}\ \frac{e^{-\frac{1}{T}(\ U(\boldsymbol{\omega}) - T\ \ln P(\mathbf{d}|\boldsymbol{\omega})\ )}}{Z} \end{aligned} \tag{12.4}$$

- falls sich der Term $\ln P(\mathbf{d}|\boldsymbol{\omega})$ durch Summation über Cliquen ausdrücken laßt, dann ist auch $P(\boldsymbol{\omega}|\mathbf{d})$ eine Gibbs-Verteilung

1. Annahme der statistischen Unabhängigkeit der Beobachtungen an den verschiedenen Stellen (unrealistisch, aber häufig getroffen), dann gilt:

$$\begin{aligned} \ln P(\mathbf{d}|\boldsymbol{\omega}) &= \ln \prod_{i=1,\ldots,N} P(d_i|\boldsymbol{\omega}) \\ &= \sum_{i=1,\ldots,N} \ln P(d_i|\boldsymbol{\omega}) \end{aligned} \quad (12.5)$$

   fasse $T \ln P(d_i|\boldsymbol{\omega})$ als zusätzliche Terme für die Clique-Potentiale $\mathcal{C}_1$ auf

$$U(\boldsymbol{\omega}|\mathbf{d}) = U(\boldsymbol{\omega}) - T \sum_{i=1,\ldots,N} \ln P(d_i|\boldsymbol{\omega}) \quad (12.6)$$

2. es gibt auch allgemeinere Bedingungen

12.2 Energieminimierung

- bei gegebener Beobachtung $\mathbf{d}$ und MRF suche:

$$\boldsymbol{\omega}^* = \underset{\boldsymbol{\omega}\in\Omega}{\operatorname{argmax}} P(\boldsymbol{\omega}|\mathbf{d}) = \underset{\boldsymbol{\omega}\in\Omega}{\operatorname{argmin}} U(\boldsymbol{\omega}|\mathbf{d})$$

- Konfigurationsraum Ω sehr groß, daher vollständige Suche nicht möglich

  1. simulated annealing (simuliertes Auskühlen)
     - zufällige Initialisierung
     - ausgehend von aktueller Konfiguration wähle zufällige eine neue (i.d.R. Änderung an einer Stelle)
     - falls diese neue Konfiguration niedrigere Energie hat, dann übernehme sie,
     - falls die Energie höher, dann übernimm sie nur mit einer gewissen Wahrscheinlichkeit,
       die im Lauf der Minimierung ab nimmt (annealing schedule)

     findet globales Optimum, falls annealing schedule genügend "langsam"

     aber: sehr aufwendig und optimaler schedule i.A. unbekannt

2. Gibbs-Sampler

   im Prinzip wie simulated annealing, aber Änderung der Konfiguration bezieht lokale Charakteristik der Gibbs-Verteilung ein

   auch sehr aufwendig

3. Iterative Conditional Modes (ICM)
   - zufällige Initialisierung
   - besuche in deterministischer Reihenfolge die Stellen
   - bei Besuch der Stelle s_i bestimme als neues Label dasjenige w_i, das lokal den maximalen Energieverlust bewirkt

   schnell, aber hängt stark von der initialen Konfiguration ab

4. Highest-Confidence-First
   - wie ICM, aber die Reihenfolge wird vom Problem abhängig gemacht
   - deterministische Initalisierung mit neuem Label “unentschieden”
   - besuche zuerst die Stellen, die aufgrund der lokalen Struktur die größte Konfidenz der (lokalen) Entscheidung versprechen

   auch schnell, führt oft zu gutem (lokalem) Optimum

12.3 Zusammenfassung

Modellierung eines Problems mit MRF:

- welche Objekte, Entitäten bilden die Stellen ?
- was ist die Label-Menge, d.h. mögliche Interpretationen
- Nachbarschaften definieren, wodurch der Kontext festlegt wird
- Clique-Potentiale definieren sowohl das a priori Wissen als die Datenabhängigkeit
- Verfahren zur Enegieminimierung auswählen

12.4 Anwendungen

12.4.1 Rekonstruktion Grauwerten mit glatten Oberflächen

Problemstellung ein Bild mit (als gestört betrachteten) Grauwerten soll rekonstruiert werden

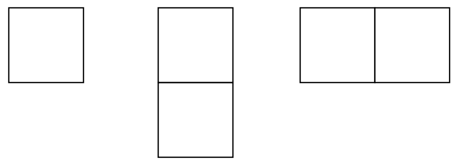
- jeder Bildpunkt bildet eine Stelle s_{ij}, die Label entsprechen dem jeweiligen Grauwert des Bildpunktes
- die beobachteten Daten sind die aufgenommenen Grauwerte $f_{ij} \hat{=} d_{ij}$
- die gesuchte Konfiguration sind die rekonstruierten Grauwerte ω_{ij}, wobei Störungen eliminiert sein sollen

Datenabhängigkeit

$$V_{\{s_{ij}\}}(\boldsymbol{\omega}) = \ln P(w_{ij} \mid \boldsymbol{\omega}) = lnP(w_{ij} \mid \omega_{ij}) = (f_{ij} - \omega_{ij})^2$$

Modellwissen

- Nachbarschaft ist die 4er-Nachbarschaft
- resultierende Cliquen:



- in der Nachbarschaft (einer Clique) sollen die Grauwerte möglichst konstant sein (glatte Oberflächen!)
  $\Rightarrow$ Clique-Potentiale

$$
\begin{aligned}
V_{\{s_{ij}, s_{kl}\}}(\boldsymbol{\omega}) &= (\omega_{ij} - \omega_{kl})^2 \\
U(\boldsymbol{\omega}) &= \sum_{\{(i,j),(k,l)\} \in \mathcal{C}_2} (\omega_{ij} - \omega_{kl})^2
\end{aligned}
$$

  (a priori Einer-Cliquen haben neutrales Potential 0)

Bemerkungen

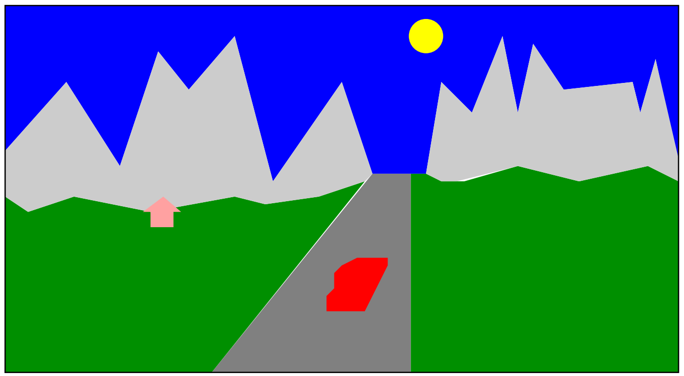
1. es werden auch andere Nachbarschaften verwendet (z.B. 8er-Nachbarschaft)
2. Erweiterungen führen einen “Kantenprozeß” in einem zweiten, gekoppelten Kanten-MRF ein

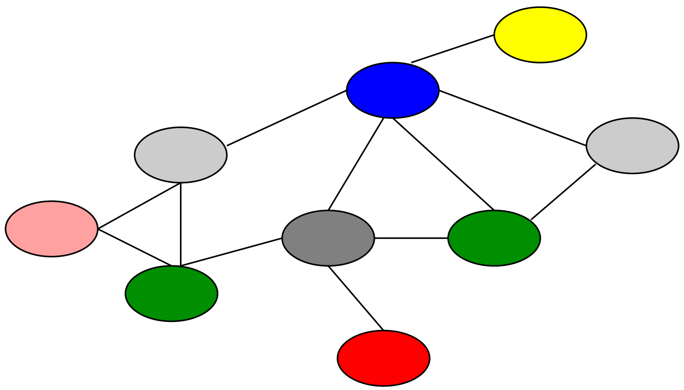
12.4.2 Klassifikation von Bildregionen

Modestino et al

Problemstellung — Klassifikation (Labelling) von segmentierten Bildregionen

Modellierung
- jede Region bildet eine Stelle s_i, die Label entsprechen den möglichen Klassen (=Interpretationen) $\mathcal{L} = \{$ Himmel, Auto, Wiese, Meer, $\ldots \}$
- die Nachbarschaft entspricht dem adjacency graph d.h. Regionen sind benachbart, gdw. sie eine gemeinsame Grenze haben





Datenabhängigkeit wiederum als Potentiale einelementiger Cliquen: quantifizierte die **lokale** Konsistenz der Merkmale der Region mit dem aktuell zugewiesenem Label, also Interpretation

$$V_{\{s_i\}}(\omega_i) = \sum_{m=1}^{M} \alpha_m \;\; B_m\left(\omega_i, X_m(s_i)\right)$$

wobei

$X_m(s_i)$	das m-te Merkmale der Region an der Stelle s_i z.B. Fläche, mittlerer Grau- und Farbwert
B_m	Basisfunktion
α_m	Parameter zur Gewichtung der Merkmale

- $B_m\left(\omega_i, X_m(s_i)\right)$ hat kleine Werte, wenn die Interpretation ω_i mit dem Merkmal X_m der Region s_i kompatibel ist
- **eine** mögliche Wahl:

$$B_m\left(\omega_i, X_m(s_i)\right) = 1 - P\left(\omega_i \mid X_m(s_i)\right)$$

wobei die a posteriori Wahrscheinlich mit Bayes aus einer markierten Stichprobe gewonnen werden kann

Modellwissen über benachbarte Regionen (i.d.R. zwei, ev. drei) anhand

- von Merkmalen der benachbarten Regionen
- der Kompatibilität der benachbarten Interpretation (z.B. Autos fliegen normal nicht im Himmel)

$$V_{\{s_i,s_j\}}(\omega_i,\omega_j) = \sum_{i=l}^{L} \beta_l \;\; C_l\left(\omega_i,\omega_j,Y_l(s_i,s_j)\right) + \beta_0 \;\; S(\omega_i,\omega_j)$$

wobei

$Y_l(s_i,s_j)$	l-tes Merkmal der Regionen an den Stellen s_i, s_j z.B. Länge der gemeinsamen Grenze, Kontrast zwischen den Regionen
C_m	Basisfunktion
β_l	Parameter zur Gewichtung der Merkmale
$S(\omega_i,\omega_j)$	Kompatibilität der benachbarten Klassen

- Wahl von C_m und S ähnlich wie bei Datenabhängigkeit, z.B. über Schätzung der a posteriori Wahrscheinlichkeiten
- Wahl der Gewichte β_l meist heuristisch

- allgemein für MRFs: das festlegen, trainieren der Parameter, d.h. im wesentlichen der Clique-Potentiale eher heuristisch